



Citation for published version:

Day, M 1998, Issues and Approaches to Preservation Metadata. in *Guidelines for digital imaging: papers given at the joint National Preservation Office and Research Libraries Group preservation conference in Warwick, 28th - 30th September 1998*. National Preservation Office, London, pp. 73-84.

Publication date:
1998

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Issues and Approaches to Preservation Metadata

Michael Day¹

UKOLN: The UK Office for Library and Information Networking,

University of Bath, Bath, BA2 7AY, United Kingdom.

<http://www.ukoln.ac.uk/>

m.day@ukoln.ac.uk

Abstract

The creation and use of metadata is likely to become an important part of all digital preservation strategies whether they are based on hardware and software conservation, emulation or migration. The UK Cedars project aims to promote awareness of the importance of digital preservation, to produce strategic frameworks for digital collection management policies and to promote methods appropriate for long-term preservation - including the creation of appropriate metadata. Preservation metadata is a specialised form of administrative metadata that can be used as a means of storing the technical information that supports the preservation of digital objects. In addition, it can be used to record migration and emulation strategies, to help ensure authenticity, to note rights management and collection management data and also will need to interact with resource discovery metadata. The Cedars project is attempting to investigate some of these issues and will provide some demonstrator systems to test them.

1. Introduction

1.1 Metadata and digital preservation

Most current discussions of metadata in the library and information communities have centred on issues of resource description and discovery (e.g. Heery, Powell and Day 1997; Dempsey and Heery 1998). Metadata is commonly understood as an amplification of traditional bibliographic cataloguing practices in an electronic environment. Perhaps the most widely known international metadata standard is the Dublin Core, an initiative that has a deliberate focus on simple resource discovery (e.g. Weibel and Hakala 1998). These are important issues. It is becoming increasingly recognised, however, that metadata has other important roles in the wider task of managing digital resources. For example, publishers and other rights owners are beginning to investigate the uses of metadata with regard to rights management (Rust 1998).

The Library of Congress Making of America II project has identified a threefold division of metadata (Making of America II 1998):

¹ Preprint of a paper published in: *Guidelines for digital imaging: papers given at the joint National Preservation Office and Research Libraries Group preservation conference in Warwick, 28th - 30th September 1998* (London: National Preservation Office, 1998), pp. 73-84.

- Descriptive metadata - primarily used for resource discovery. Formats currently used include the MARC formats, Encoded Archival Description (EAD) and Dublin Core.
- Structural metadata - data that a system can use to help present a particular digital object to a user.
- Administrative metadata - data that allows the management of a digital collection.

Preservation is essentially about management. In this scheme, preservation metadata (as with rights metadata) is a specialised form of administrative metadata.

1.2 The Cedars project

The Cedars (CURL exemplars in digital archives) project is funded by the Joint Information Systems Committee (JISC) of the UK higher education funding councils under Phase III of its Electronic Libraries (eLib) Programme. The project is administered through the Consortium of University Research Libraries (CURL) with lead sites based at the Universities of Cambridge, Leeds and Oxford.

Cedars is a project that aims to address strategic, methodological and practical issues relating to digital preservation (Day 1998a). A key outcome of the project will be to improve awareness of digital preservation issues, especially within the UK higher education sector. Attempts will be made to identify and disseminate:

- Strategies for collection management
- Strategies for long-term preservation

These strategies will need to be appropriate to a variety of resources in library collections. The project will also include the development of demonstrators to test the technical and organisational feasibility of the chosen preservation strategies. One strand of this work relates to the identification of preservation metadata and a metadata implementation that can be tested in the demonstrators.

1.3 Digital imaging technology and preservation

The use of digital imaging technology by libraries, archives and museums is largely concerned with the creation of digital surrogates of analogue material and tends to be motivated by two interrelated concerns:

- Improving access to information; and
- Preservation - ensuring continuing access to this information.

With analogue materials there is a potential conflict of interest between improving access and ensuring long-term preservation. The use of digital imaging technology to create surrogates of analogue material, however, broadens the preservation strategies that can be adopted by custodial organisations (Weber, H. and Dörr 1997). Once an information object has been digitally imaged - assuming that certain minimum quality standards apply - it would be possible to output the

image on a preservation-quality computer output microfilm (COM) for long-term retention while at the same time maintaining digital versions of the same for access. In this way, custodial organisations can separate the medium used for preservation from the media used for production and use (Kenney and Conway 1994, p. 19).

Even so, it is likely that some custodial organisations will develop collection management strategies that recommend the long-term retention of some digitally imaged material. Once digitised images have been created and there is a specific requirement for their long-term preservation, much the same preservation considerations will apply to them as to information objects that are 'born digital'.

The Cedars project will investigate some of the issues of preserving this type of material. The project demonstrators will include digitally imaged information objects like the journals made available through the ILEJ (Internet Library of Early Journals) project (Jupp 1997) and digital images of manuscript fragments held in the Taylor-Schechter Genizah Collection at Cambridge University Library (Taylor-Schechter Genizah Research Unit 1998). The Cedars project scope, however, is mostly concerned with digital resources that are 'born digital'. With these, interpretation of the resource is often fixed to its existence as a digital object and a human-readable surrogate will not always be adequate to express this. This is why the digital preservation of information created digitally is extremely important.

2. Digital preservation and metadata

2.1 Preservation metadata for digital preservation strategies

Digital preservation is as much a strategic problem as a technical one. For this reason it is imperative that the strategic context for the creation and preservation of digital resources be taken into account. This process has been eased by the appearance of an UK Arts and Humanities Data Service (AHDS) report that outlines a policy framework applicable to the three main stages in the life cycle of a digital resource: creation, management/preservation and use (Beagrie and Greenstein 1998). Solving the technical issues of digital preservation will be important but is essentially subordinate to these wider, strategic, considerations.

The main technical problems of digital preservation relate to inadequate media longevity, rapid hardware obsolescence and dependencies on particular software products. There are currently three main approaches to digital preservation (Ross 1997). The metadata issues raised will differ according to which particular strategy is adopted but it should be noted that metadata strategies have an important part to play in all three.

2.1.1 Technology preservation

This approach proposes that digital data should be preserved on a stable medium (and 'refreshed' or copied to new media as necessary) and associated with preserved copies of the original application software, the operating system that this would normally run under and the relevant hardware platform. This strategy may have some value for particularly important (or historic) examples of software or hardware or could be useful for the museum community (Swade 1992) but in the long-term is likely to be expensive and impractical. Tony Hendley (1998, p. 17)

comments that the technology preservation approach "cannot be regarded as viable for anything other than the short to medium term". He also comments that it could be used as "a relatively desperate measure in cases where valuable digital resources cannot be converted into hardware and/or software independent formats and migrated forwards".

2.2 Emulation

The second main suggested approach to digital preservation is technology emulation. This strategy relies (as with technology preservation) on the preservation of the original data in its original format. Instead of preserving the host software and hardware, software engineers would build emulator programs that would mimic the behaviour of obsolete hardware platforms and emulate the relevant operating system (Rothenberg 1995). In practice, data could be encapsulated together with the application software used to create it and a description of the required hardware and software environment. To facilitate future use, Jeff Rothenberg (1996) suggests attaching 'annotation metadata' to the surface of each encapsulation which would both "explain how to decode the obsolete records contained inside the encapsulation and to provide whatever contextual information is desired about these records".

Emulation is an important strategy that has potential applications where the look and feel of an original digital resource is of importance but where it is not worth investing in expensive technology preservation. Hendley (1998, p. 18), however cautions against relying solely on this approach and comments that collection managers "would be depending on the technical ability of the software engineers to emulate a specific environment and sustain it".

A related approach is the Digital Rosetta Stone (DRS) model developed by Steven Robertson of the United States Air Force (Heminger and Robertson 1998). In this model, digital documents would be maintained in their original file formats. In conjunction with this, a 'metaknowledge archive' (MKA) would be created to store the "the vast amounts of knowledge needed to recover digital data from a superseded media and to reconstruct digital documents from their original formats" (Robertson 1996, p. 23). The MKA would be a collection of the knowledge and processes necessary to recover and reconstruct digital documents maintained in their original file formats. This data would be used to re-create (or emulate) the hardware and software necessary to recover data from obsolete storage media and reconstruct the digital documents. The DRS model, like technology preservation strategies, might have an application as a backup strategy where other approaches have failed. As a long-term solution to digital preservation, however, it is likely to be expensive.

2.3 Migration

A third approach to digital preservation is the periodic migration of digital information from one hardware or software environment to another. The Task Force on the Archiving of Digital Information (1996) has produced a good (much cited) definition:

Migration is the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology.

The point of migration is to transfer to new formats while, where possible, preserving the integrity of the information. A digital archive could convert incoming digital objects into a small number of 'standard' formats. For example textual data could be stored in a relatively software independent format like ASCII, in widely used proprietary formats like the Portable Document Format (PDF) or in formats based on applications of SGML (Coleman and Willis 1997). Over time, data would be copied and refreshed as necessary and periodically migrated to new formats for use with new generations of hardware and software.

Metadata has an important role in any successful migration strategy. Such a strategy will depend upon metadata being created to record the migration history of a digital object. In addition there is a need for contextual information to be recorded (and preserved) so that a future user can understand the technological environment in which a particular digital object was created. David Bearman (1994, p. 302) says that "content, structure and context information must be linked to software functionality that preserves their executable connections or representations of their relations must enable humans to reconstruct the relations that pertained in the original software environment".

2.2 Metadata and authentication

In addition to the purely technical problems of digital preservation, there will be a need to address problems of what Peter Graham (1994) calls intellectual preservation. How will users know that the digital object that they retrieve is the one that they want? How will administrators of digital repositories know that their holdings have not been subject to unauthorised changes, either accidental or deliberate?

The use of persistent and unique digital identifiers has a potentially important role in this regard. New identifiers would need to be assigned each time a particular digital object is updated or migrated. Current digital identifier initiatives include the Uniform Resource Name (URN) which is being developed for the Internet community by working groups of the Internet Engineering Task Force (Sollins and Masinter 1994) and the Digital Object Identifier (DOI), an initiative of the Association of American Publishers (e.g. Bide 1998). Legacy identifiers will also continue to be used for some of the digital objects that will need preservation, so - for example - some publishers could assign International Standard Book Numbers (ISBNs) to CD-ROMs or generate Serial Item and Contribution Numbers (SICIs) for online journal articles. On the other hand, other items in the project scope, electronic ephemera like Web pages or example, are unlikely to be assigned digital identifiers except for Uniform Resource Locators (URLs).

An additional approach to ensuring the authenticity of a given digital object would be to use simple cryptographic techniques like the production of a validation key value or checksum for each resource in a digital archive. An authentication checksum could be computed from each resource in a digital archive and stored with the descriptive metadata. When a user, or the archive, wants to retrieve the resource at a later date this checksum could be computed again and compared with the checksum recorded in the metadata. If the two agree there can be confidence that the document retrieved is the one referred to by the descriptive metadata. This general approach has been adopted for use by the European Telematics for Libraries project BIBLINK (Peacock and Powell 1998). Other possible approaches to the problem could use other cryptographic techniques like digital signatures. The World Wide Web Consortium (W3C)

Digital Signature Working Group (DSig), for example, has developed digital signatures - currently implemented in PICS (Platform for Internet Content Selection) technology - for making assertions about particular Web information resources.

It is worth noting in addition that archivists and records managers share these professional concerns with preserving the authenticity and integrity of digital objects (e.g. Duff 1995; Duranti and MacNeil 1995). The University of Pittsburgh Electronic Records Project, for example, has defined a metadata model for business-acceptable communications that emphasises the preservation of a record's 'evidentiality' (Bearman and Sochats 1996).

2.3. Metadata and rights management

Solving rights management issues will be vital in any digital preservation programme. Typically, custodial organisations do not have physical custody of digital objects created or made available by other stakeholders (e.g. authors or publishers). Instead they will negotiate rights to this information for a specific period of time. Permissions to preserve digital information objects will also need to be negotiated with rights holders and any such agreement may, or may not, permit end user access. A digital archive will have to collect and store any relevant rights management information which could be stored as part of the descriptive metadata. This could also be used to manage access.

The Cedars project has a Content Issues Working Group that will negotiate with rights holders for the use of material in the demonstrators. This group will also examine broader issues relating to long term storage and access to copyright material and work closely with rights owners to recommend best practice.

2.4 Other metadata issues

2.4.1 Resource discovery

There will be no point preserving large amounts of digital information unless there is some consideration of resource discovery issues. Information objects that have been digitally preserved will need descriptive metadata that can aid resource discovery or ideally that can interact with other resource discovery systems, including existing library catalogues.

Recommendations on relevant resource discovery formats (e.g. Dublin Core, MARC) and metadata frameworks like the Resource Description Format (RDF) will constitute an important part of Cedars work on metadata.

2.4.2 Collection Management

There will be no need to preserve all existing digital resources, as not all will be worthy of long-term preservation. The Cedars project is interested in helping to develop suitable collection management policies for research libraries. This work could build on work carried out on selection criteria for Internet subject gateways produced by the EU funded DESIRE project (Hofman, *et al.* 1997). The existence, or otherwise, of appropriate metadata for preservation, resource discovery and other purposes will be essential to allow appropriate decisions to be made about what items need to be included in digital collections and how these should be

administered.

2.4.3 Metadata management

Another important series of issues relate to the management and migration of any proposed preservation metadata system. For example, metadata can either be stored in a database and linked (in some way) to the original resource or embedded in (or otherwise directly associated with) the original resource. Resource discovery and rights management metadata could form part of a searchable database that gives access to digital objects, while metadata specifying the technical formats used, the migration strategies operated and a document's use history could be stored closer to the document itself. Over time, this metadata will itself have to be subject to migration and authentication strategies.

3. Current initiatives and data models

The Cedars Access Issues Working Group has produced a preliminary study of preservation metadata and the issues that surround it (Day 1998b). This study describes some digital preservation initiatives and models with relation to the Cedars project and will be used as a basis for the development of a preservation metadata implementation in the project. The remainder of this paper will describe some of the metadata approaches found in these initiatives.

3.1 *The RLG Working Group on Preservation Issues of Metadata*

The Research Libraries Group constituted a Working Group on the Preservation Issues of Metadata in May 1997 and its final report (RLG Working Group 1998) is perhaps the best current assessment of the preservation and metadata requirements of digital imaging technology. The working group limited itself to a consideration of the data elements that describe digital image files, arguing that other specialist groups could be constituted to analyse other formats when the need becomes more pressing. The group also examined two 'core' metadata formats, the Dublin Core and the Program for Co-operative Cataloging's USMARC-based core record standard, so that the group could specify the metadata elements extra to these core element lists that would be important to serve preservation needs. The sixteen metadata elements deemed crucial for the continued viability of a digital master file were:

Element	Brief description
Date	Date file is created
Transcriber	Name of agency (or individual) responsible for transcribing the metadata
Producer	Agency (or individual) responsible for the physical creation of the file.
Capture Device	Make and model of digital camera or scanner.

Capture Details	<ol style="list-style-type: none"> 1. Name of scanner software, version information, scanner settings, gamma correction, etc. 2. Digital camera lens type, focal length, light source type, etc.
Change History	A record of modifications made to the file.
Validation Key	A mechanism allowing one to verify that the electronically transmitted file is what it purports to be.
Encryption	The technique by which data is encryption before transmission.
Watermark	Indicates whether (or not) some bits in the file have been altered in order to create a digital fingerprint or similar.
Resolution	Resolution determined by pixel dimensions, pixels per inch or dots per inch.
Compression	Indicate whether (or not) file has been conversed.
Source	Physical characteristics of the source, etc.
Color	Pixel depth.
Color Management	Systems (if any) used to improve consistency of colour.
Color Bar/Gray Scale Bar	Indicates presence (or not) of either, with type.
Control Targets	Information about targets included in scanned file.

The RLG Working Group also published three potential implementations of these metadata elements for discussion and experimentation: firstly a hypothetical Dublin Core record, secondly a mapping of the elements to USMARC and thirdly a simple XML implementation. The RLG working group report gives a useful indication of some of the individual metadata elements that need to be captured to help ensure some degree of digital preservation. The report encourages institutions to implement the RLG element set and to share their efforts with the rest of the community.

Other relevant metadata elements for are identified in the Making of America II Testbed Project White Paper, which include elements recorded at the point of capture for a digital master images, context metadata and rights management information (Making of America II 1988).

3.2 Open Archival Information System (OAIS)

The Cedars study suggests that there might be some value in adopting (or adapting) relevant metadata models. The most important existing model is the Reference Model for an Open Archival Information System (OAIS) published by the Consultative Committee for Space Data

Systems (CCSDS). OAIS is an ISO initiative (co-ordinated by the CCSDS) that defines a high-level reference model for archives originally concerned with the long-term preservation of digital information obtained from observations of terrestrial and space environments but which would be applicable to other long-term digital archives. An archive (in OAIS terms) consists of "an organisation of people and systems, that has accepted the responsibility to preserve information and make it available for one or more designated communities" (CCSDS 1998). The OAIS model has a 'taxonomy of archival information object classes' (CCSDS 1998, pp. 50-57) that includes:

Content Information:	This is the information that is the primary object of preservation. This contains the primary Digital Object and Representation Information needed to transform this object into meaningful information.
Preservation Description Information:	<p>This would include any information necessary to adequately preserve the Content Information with which it is associated. It includes:</p> <ul style="list-style-type: none">• Reference Information - (e.g. identifiers),• Context Information (e.g. subject classifications),• Provenance Information (e.g. copyright)• Fixity Information (that documents the authentication mechanisms).
Packaging Information:	The information that binds and relates the components of a package into an identifiable entity on a specific media.
Descriptive Information:	The information that allows the creation of Access Aids - to help locate, analyse, retrieve or order information from an OAIS.

This taxonomy includes (and refines) many of the metadata types discussed in the Cedars report. Any high-level architecture developed for Cedars will probably conform to the OAIS model.

3.3 National Library of Australia PANDORA logical data model

A separate model is the 'logical data model' developed by the National Library of Australia for its Preserving and Accessing Networked DOcumentary Resources of Australia (PANDORA) project (National Library of Australia 1997). This model is based on an entity-relationship diagram that identifies the logical entities that need to be supported by the PANDORA system. The highest level entities are:

- Identification
- Selection and negotiation

- Capture
- Preservation
- Rights Management and Access Control

Each of these is divided into further entities and each of these into metadata attributes.

Preservation metadata is defined as "entities required to support the management of copies within the archive, including activities to ensure both the immediate and long term accessibility of the item". The entities include 'File' and 'File Type' (e.g. M/S Word, HTML, ASCII, JPEG, PDF, TIFF, etc.), 'Format' and 'Format Type' (e.g. Online, Diskette, CD-ROM, etc.). The notes on Format suggest that such information should be recorded at the selection stage as part of technical assessment. It also recommends that "a history trail is kept of the format of the copy at the time of archiving and any technical processing that has been conducted on the copy to ensure preservation and access".

A copy of a publication may be converted from one format to another to improve accessibility in the host environment or to help migrate whole categories of publication to a new technology base. Generally a conversion from one format to another will involve tangible formats (e.g., to transfer files from diskette to CD-R) but there may also be a requirement to convert data from a tangible to online format or vice versa. When a format is converted to another format type, a record will be maintained of the conversion process, with a link to the new format type (National Library of Australia 1997).

3.4 Resource Description Framework

The Cedars project will not just be adopting (or adapting) a high-level data model like OAIS. It will attempt to develop demonstrators that will implement selected aspects of digital preservation including those related to metadata. The precise nature of the metadata implementation has yet to be decided by the project but the Resource Description Framework (RDF) being developed under the auspices of the World Wide Web Consortium (W3C) is of potential interest. RDF provides a data model for describing resources and proposes an Extensible Markup Language (XML) based syntax based on this data model (World Wide Web Consortium 1998). The need to aggregate multiple sets of metadata was noted at the second Dublin Core workshop and was the principle that underlay the formulation of the Warwick Framework container architecture (Lagoze, Lynch and Daniel 1996; Weibel and Lagoze 1997). Similarly, RDF aims to facilitate modular interoperability among different metadata element sets by creating what Eric Miller (1998) calls "an infrastructure that will support the combination of distributed attribute registries". The modular principle of RDF means that Cedars-defined preservation metadata elements could be aggregated with metadata types defined for other purposes, e.g. Dublin Core for simple resource discovery or structured data about terms and conditions. This type of interoperability is likely to be a useful aspect of preservation metadata systems.

4. Conclusions

The definition and implementation of preservation metadata systems is going to be an important part of the work of custodial organisations in the digital environment. Projects like Cedars are attempting to investigate some of the relevant issues and provide some demonstrator systems that

can test them. Individuals and organisations interested in the long-term preservation of digital information need to note of preservation metadata issues. The future of our digital collections will depend, to some extent, on how carefully we respond to this challenge.

5. References

- Beagrie, N. and Greenstein, D., 1998, *A Strategic Policy Framework for Creating and Preserving Digital Collections*. London: Arts and Humanities Data Service, 14 July.
<URL:<http://ahds.ac.uk/manage/framework.htm>>
- Bearman, D., 1994, *Electronic evidence: strategies for managing records in contemporary organizations*. Pittsburgh, Penn.: Archives and Museum Informatics.
- Bearman, D. and Sochats, K., 1996, *Metadata requirements for evidence*. Pittsburgh, Penn.: University of Pittsburgh, School of Information Science.
<URL:<http://www.lis.pitt.edu/~nhprc/BACartic.html>>
- Bide, M., 1998, *In search of the Unicorn: the Digital Object Identifier from a user perspective*, rev. ed. British National Bibliography Research Fund Report 89. London: Book Industry Communication. <URL:<http://www.bic.org.uk/bic/unicorn2.pdf>>
- Coleman, J. and Willis, D., 1997, *SGML as a framework for digital preservation and access*. Washington, D.C.: Commission on Preservation and Access.
- Consultative Committee for Space Data Systems, 1998, *Reference Model for an Open Archival Information System (OAIS)*, ed. L. Reich and D. Sawyer. CCSDS 650.0-W-4.0. White Book, Issue 4, 17 September. Latest version available from:
<URL:http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html>
- Day, M.W., 1998a, CEDARS, digital preservation and metadata. In: *Sixth DELOS Workshop: Preservation of Digital Information, Tomar, Portugal, 17-19 June 1998*. ERCIM-98-W003. Le Chesnay: European Research Consortium for Informatics and Mathematics, pp. 53-58.
<URL:<http://www.ercim.org/publication/ws-proceedings/DELOS6/>>
- Day, M.W., 1998b, *Metadata for Preservation*. CEDARS Project Document AIW01.
<URL:<http://users.ox.ac.uk/~cedars/Papers/aiw01.html>>
- Dempsey, L. and Heery, R., 1998, Metadata: a current view of practice and issues. *Journal of Documentation*, 54 (2), pp. 145-172.
- Duff, W., 1995, Ensuring the preservation of reliable evidence: a research project funded by the NHPRC. *Archivaria*, 42, pp. 28-45.
- Duranti, L. and MacNeil, H., 1995, The protection of the integrity of electronic records: an overview of the UBC-MAS Research Project. *Archivaria*, 42, pp. 46-67.
- Graham, P.S., 1994, Long-term intellectual preservation. In *Digital imaging technology for preservation*, ed. N.E. Elkington, Mountain View, Calif.: Research Libraries Group, pp. 41-57.
- Heery, R., Powell, A. and Day, M., 1997, *Metadata*. Library and Information Briefings, 75.

London: South Bank University, Library Information Technology Centre.

Heminger, A.R. and Robertson, S.B., 1998, Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents. In: *Sixth DELOS Workshop: Preservation of Digital Information*. ERCIM-98-W003. Le Chesnay: European Research Consortium for Informatics and Mathematics, 35-43. <URL:<http://www.ercim.org/publication/ws-proceedings/DELOS6/>>

Hendley, T., 1998, *Comparison of methods & costs of digital preservation*, British Library Research and Innovation Report, 106. London: British Library and Innovation Centre.

Hofman, P., Worsfold, E., Hiom, D., Day, M. and Oehler, A., 1997, *Specification for resource description methods: 2, Selection criteria for quality controlled information gateways*. DESIRE: Development of a European Service for Information on Research and Education, Deliverable 3.2 (2). <URL:<http://www.ukoln.ac.uk/metadata/desire/quality/>>

Jupp, B., 1998, The Internet Library of Early Journals. *Aslib Proceedings*, 49 (6), 153-158. <URL:<http://www.bodley.ox.ac.uk/ilej/papers/paper01.htm>>

Kenney, A. and Conway, P., 1994, From analog to digital: extending the preservation tool kit. In *Digital imaging technology for preservation*, ed. N.E. Elkington. Mountain View, Calif.: Research Libraries Group, pp. 11-24.

Lagoze, C., Lynch, C.A. and Daniel, R., 1996, *The Warwick Framework: a container architecture for aggregating sets of metadata*. Cornell Computer Science Technical Report TR96-1593. <URL:<http://cs-tr.cs.cornell.edu:80/Dienst/UI/1.0/Display/ncstrl.cornell/TR96-1593/>>

Making of America II, 1998, *The Making of America II testbed project white paper*. Version 1.03, March 16. <URL:http://sunsite.berkeley.edu/MOA2/wp-v1_03.html>

Miller, E., 1998, An introduction to the Resource Description Framework. *D-Lib Magazine*, May. <URL:<http://www.dlib.org/dlib/may98/miller/05miller.html>>

National Library of Australia, 1997, *PANDORA Logical Data Model*, Version 2, 10 November. <URL:<http://www.nla.gov.au/pandora/ldmv2.html>>

Peacock, I. and Powell, A., 1998, BIBLINK.Checksum - an MD5 message digest for Web pages. *Ariadne* (Web version), no. 17, September. <URL:<http://www.ariadne.ac.uk/issue17/biblink/>>

RLG Working Group on Preservation Issues of Metadata, 1998, *Final report*. Mountain View, Calif.: Research Libraries Group, May. <URL:<http://www.rlg.org/preserv/presmeta.html>>

Robertson, S.B., 1996, *Digital Rosetta Stone: a conceptual model for maintaining long-term access to digital documents*. Thesis (MSc), Air Force Institute of Technology, Graduate School of Logistics and Acquisition Management. <URL:http://www.au.af.mil/au/database/research/ay1996/afit_la/rober_sb.htm>

Ross, S., 1997, Consensus, communication and collaboration: fostering multidisciplinary co-operation in electronic records. In: *Proceedings of the DLM-Forum on Electronic Records*,

Brussels, 18-20 December 1996. INSAR: European Archives News, Supplement II. Luxembourg: Office for Official Publications of the European Communities, pp. 330-336.

Rothenberg, J., 1995, Ensuring the longevity of digital documents. *Scientific American*, 272 (1), pp. 24-29.

Rothenberg, J., 1996, *Metadata to support data quality and longevity*. Proceedings of the 1st IEEE Metadata Conference, NOAA Complex, Silver Spring, Md., 16-18 April.
<URL:http://www.computer.org/conferen/meta96/rothenberg_paper/ieee.data-quality.html>

Rust, G., 1998, Metadata: the right approach. An integrated model for descriptive and rights metadata in e-commerce. *D-Lib Magazine*, July/August.
<URL:<http://www.dlib.org/dlib/july98/rust/07rust.html>>

Sollins, K. and Masinter, L., 1994, *Functional Requirements for Uniform Resource Names*. RFC 1737. <URL:<http://ds.internic.net/rfc/rfc1737.txt>>

Swade, D., 1992, The problems of software conservation. *History and Computing*, 4 (2).
<URL:http://www.cs.man.ac.uk/CCS/simulate/sim_home.htm>

Task Force on the Archiving of Digital Information, 1996, *Preserving digital information: report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group*. Washington, D.C.: Commission on Preservation and Access. <URL:<http://www.rlg.org/ArchTF/>>

Taylor-Schechter Genizah Research Unit, 1998, *The Genizah On-Line Database (GOLD)*. Cambridge: Cambridge University Library. <URL:<http://www.lib.cam.ac.uk/Taylor-Schechter/GOLD/>>

Weber, H. and Dörr, M., 1997, *Digitisation as a method of preservation?* Amsterdam: European Commission on Preservation and Access.

Weibel, S. and Hakala, J., 1998, DC-5: The Helsinki Metadata Workshop: a report on the workshop and subsequent developments. *D-Lib Magazine*, February.
<URL:<http://www.dlib.org/dlib/february98/02weibel.html>>

Weibel, S.L. and Lagoze, C., 1997, An element set to support resource discovery: the state of the Dublin Core, January 1997. *International Journal on Digital Libraries*, 1(2), pp. 176-186.

World Wide Web Consortium, 1998, *Resource Description Framework (RDF) model and syntax specification*, eds. O. Lassila and R. Swick. W3C Working Draft.
<URL:<http://www.w3.org/TR/WD-rdf-syntax/>>

6. Acknowledgements

UKOLN is funded by the British Library Research and Innovation Centre (BLRIC), the Joint Information Systems Committee (JISC) of the UK higher education councils, as well as by project funding from several sources. UKOLN also receives support from the University of Bath, where it is based. The views expressed in this paper do not necessarily reflect those of the Cedars project, UKOLN or their funding bodies.

The author would like to thank Kelly Russell (Cedars Project Manager) and Dr Mark Nicholls (Deputy Keeper of Manuscripts, Cambridge University Library) for comments on an earlier draft of this paper.

For more information on the Cedars project, please contact: Kelly Russell, Cedars Project Manager, Edward Boyle Library, University of Leeds, Leeds LS2 9JT. <k.l.russell@leeds.ac.uk> <URL:<http://www.leeds.ac.uk/cedars/>>.